Spontaneous virulence loss in natural populations of Listeria monocytogenes

Spontaneous virulence loss in natural populations of *Listeria monocytogenes*

Hélène Bracq-Dieye[a,b], Viviane Chenal-Francisque[a,b]*, Mylène M. Maury[a,b], Lei Han[c], Guillaume Vales[a,b], Alexandre Leclercq[a,b], Alexandra Moura[b,d], Mariela Scortti[c,e], Olivier Disson[b,d], Edith Gouin[f], José A. Vázquez-Boland[c,e#], Marc Lecuit[a,b,d,g#]

[a] Institut Pasteur, National Reference Centre and WHO Collaborating Centre for *Listeria*, Paris, France

[b] Institut Pasteur, Biology of Infection Unit, Paris, France

[c] Microbial Pathogenesis Unit, Medical School (Biomedical Sciences), University of Edinburgh, Edinburgh, UK

[d] Inserm U1117, Paris, France

[e] Division of Infection & Immunity, The Roslin Institute, University of Edinburgh, Edinburgh, UK.

[f] Institut Pasteur, Bacteria-Cell Interactions Unit, Paris, France

[g] Paris Descartes University, Sorbonne Paris Cité, Institut Imagine, Necker-Enfants Malades University Hospital, Division of Infectious Diseases and Tropical Medicine, APHP, Paris, France

ML and JVB share senior authorship. Address correspondence to marc.lecuit@pasteur.fr and v.boland@ed.ac.uk

HBD, VCF, MMM and LH contributed equally to this work.

* Present address: Viviane Chenal-Francisque, Biology of Intracellular Bacteria, Institut Pasteur, Paris, France

Running Title: Spontaneous loss of *Listeria monocytogenes* virulence

Keywords: *Listeria monocytogenes*, virulence, hemolysis

Word counts: Abstract = 227 words; importance = 128 words; text = 7,335 words.

Abstract

Listeria monocytogenes (*Lm*) pathogenesis depends on its ability to escape from the phagosome of the host cells via the action of the pore-forming toxin listeriolysin O (LLO). Expression of the LLO-encoding gene (*hly*) requires the transcriptional activator PrfA, and both *hly* and *prfA* genes are essential for *Lm* virulence. Here we used the hemolytic activity of LLO as a phenotypic marker to screen for spontaneous virulence-attenuating mutations in *Lm*. Sixty (0.1%) non-hemolytic isolates were identified among a collection of 57,820 confirmed *Lm* strains isolated from a variety of sources. In most cases (56/60), the non-hemolytic phenotype resulted from nonsense, missense or frameshift mutations in *prfA*. Five strains carried *hly* mutations leading to a single amino acid substitution (G299V) or a premature stop codon causing strong virulence attenuation in mice. In one strain, both *hly* and *gshF* (encoding a glutathione synthase required for full PrfA activity) were missing due to genomic rearrangements likely caused by a transposable element. The PrfA/LLO loss-of-function mutants belonged to phylogenetically diverse clades of *Lm* and most were identified among non-clinical strains (57/60). In line with the extremely low frequency of loss of virulence mutations, we show that *prfA* and *hly* are under purifying selection. Although occurring at a low frequency, $PrfA^{-}/LLO^{-}$ mutational events in *Lm* lead to niche restriction and open an evolutionary path for obligate saprophytism in this facultative intracellular pathogen.

Importance

The hemolytic phenotype of *Lm* is a key identification criterion in food and clinical microbiology. Here we characterized 60 non-hemolytic *Lm* strains, identified by screening a vast collection of natural *Lm* isolates collected in the context of epidemiological surveillance of listeriosis. Phenotypic and genomic analyses demonstrated that the absence of hemolysis was due to loss-of-function mutations in *prfA* or *hly*, leading to strong virulence attenuation in

mice. We also identified the first natural *Lm* strain which spontaneously lost the *gshF* gene, required for the PrfA-dependent transcriptional activation of *hly* and other virulence genes. Previous phylogenomic studies have indicated that some non-pathogenic *Listeria* species derive from pathogenic ones, and the virulence-attenuating mutations characterized in this study illustrate the possible early events that could have determined their emergence and evolution.

Introduction

Listeria monocytogenes (*Lm*) is a foodborne pathogen that can cause a severe invasive disease in people and animals, called listeriosis. As a facultative intracellular bacterium, *Lm* has evolved a range of virulence determinants allowing intracellular survival (1, 2). One key virulence factor is listeriolysin O (LLO), a pore-forming toxin responsible for the characteristic ß-hemolytic phenotype of *Lm* that allows the bacterium to escape from the phagosome of host cells and replicate intracellularly (3, 4). LLO is encoded by *hly*, located in the *Listeria* Pathogenicity Island 1 (LIPI-1) (5). Expression of the genes within this central pathogenicity locus, including *hly*, is under the control of the transcriptional activator PrfA, the master regulator of *Lm* virulence genes (6, 7).

The hemolytic activity conferred by LLO is considered a cardinal marker for *Lm* detection and/or identification in clinical and food microbiology. *Lm* is divided into four phylogenetic lineages (8-10), 13 serotypes (11) that can be approximated by PCR serogrouping (12), and more than 100 clonal complexes (CCs, as defined by multilocus sequence typing (MLST)) (13), which are unevenly virulent (14). Weakly or non-hemolytic *Lm* strains have been reported (15-19), but the frequency and phylogenetic diversity of the strains displaying an altered hemolysis phenotype is unknown, as well as their underlying genetic and microbiological features.

This study aimed at (*i*) estimating the frequency of naturally-occurring non-hemolytic *Lm* isolates and their distribution among *Lm* lineages and MLST clonal complexes, (*ii*) understanding the molecular bases of the non-hemolytic phenotype and (*iii*) assessing its impact on virulence. By using phenotypic and genomic approaches, mutagenesis and *in vivo* assays, we show that mutations leading to loss of hemolytic activity in *Lm*, although rare, affect a wide range of clonal complexes of the major lineages I and II and lead to a decreased virulence.

Results

Identification and characterization of non-hemolytic *Lm* strains

We examined the prevalence of non-hemolytic *Lm* strains among the 57,820 *Lm* isolates collected between 1987 and 2008 at the French National Reference Centre for *Listeria* (NRCL) and the WHO Collaborating Centre for *Listeria* (WHOCCL). Sixty *Lm* isolates (0.1%) were identified as non-hemolytic on horse blood agar plates. These were isolated from food (n = 33), food production environments (n = 2), non-human unknown sources (n = 22) and human clinical cases (n = 3). Phenotypic characterization using the API *Listeria* system confirmed all 60 non-hemolytic isolates as *Lm*. These belonged to lineages I (n = 23, 38.3%) and II (n = 37, 61.7%) and were grouped within serogroups IIa (n = 36), IVb (n = 13), IIb (n = 10) and IIc (n = 1) (**Table S1**). MLST showed that the 60 non-hemolytic isolates belonged to 15 different clonal complexes, including the "hypovirulent" CC9 (n = 1), CC121 (n = 3), CC31 (n = 20) and ST13 (n = 3) (14, 20), but also the "hypervirulent" CC1 (n = 3), CC2 (n = 7), CC4 (n = 1) and CC6 (n = 1) (14) (**Table S1** and **Fig. 1**). Core genome MLST (cgMLST) typing identified 39 different cgMLST types (CTs) (21). Nine CTs comprised more than one strain, suggesting a possible epidemiological link between them (21) (**Table S1**). In particular, among the twenty non-hemolytic CC31 strains, ten belonged to CT878 and two

belonged to CT2659, suggesting that the overrepresentation of CC31 could be in part due to multiple sampling of the same source in the context of an epidemiological investigation. These results show that non-hemolytic strains are phylogenetically very diverse and that the loss of hemolytic activity is caused by independent events across the *Lm* population.

To investigate the impact of the loss of hemolytic activity in *Lm* fitness, we analyzed the growth of all non-hemolytic strains in BHI at 22°C and 37°C, using EGDe as control (**Fig. S1**). At 22°C, in a large majority of cases, the growth of non-hemolytic strains was within the same range as EGDe, as revealed by the areas under their growth curves (AUCs). On the contrary, at 37°C, temperature in which *prfA* is known to be maximally expressed (22), most of the non-hemolytic strains showed lower growth (lower AUCs) than EGDe. Some of the non-hemolytic strains showed particularly decreased fitness in one or both temperature: CLIP 2000/86467 ($PrfA_{T170*}$, at 22°C), CLIP 1998/75799 ($PrfA_{I51*}$-LLO_{N261*}, at 37°C) and, at both temperatures, strains CLIP 1998/76801 (Δ*hly*-Δ*gshF*), CLIP 1996/70991 ($PrfA_{Q21*}$), CLIP 1994/58618 ($PrfA_{A129P}$) and CLIP 1996/71614 ($PrfA_{Y207*}$) (**Fig. S1**).

Molecular basis of non-hemolytic phenotype – PrfA variants and activity

The central regulator of *Listeria* virulence, PrfA, is required for the expression of a set of key virulence determinants, known as the PrfA regulon, including the *hly* gene (6, 7, 23). Consequently, mutations altering the function of either PrfA or LLO could lead to a non-hemolytic phenotype. Sequence analyzes identified frameshifts and missense and nonsense mutations in *prfA* in 56 non-hemolytic strains, leading to amino-acid substitutions or protein truncations in PrfA (**Fig. 1**; **Table S1**). Phenotypic analysis in PrfA-activating and non-activating conditions using the PrfA-dependent virulence factors PlcB (phospholipase C) and Hpt as reporters (see Materials and Methods) (24) confirmed the complete loss of function of the central virulence gene regulator in all of these strains (**Fig. 1; Fig. S2**).

Forty-three out of the 56 PrfA$^-$ strains, distributed in lineages I and II, expressed a truncated PrfA at 14 distinct positions distributed along the entire PrfA protein (**Table S1)**. All analyzed strains of CC59 and CC31 exhibited a truncation at positions 59 and 185, respectively, suggesting a common ancestor for each of these groups of strains. Seven PrfA$^-$ strains presented a single amino-acid substitution in PrfA as compared to the reference strain EGDe (accession number: NC_003210). Among them, one occurred in the β-roll region of PrfA (G72D, strain CLIP 1997/75561, CC9). Mutations located in this region are known to affect PrfA activation or the ability of PrfA to form a stable complex with the RNA polymerase and initiate transcription of the target virulence genes (25-27). One PrfA$^-$ mutation occurred in the DNA-binding helix-turn-helix (HTH) domain of PrfA (G175C, strain CLIP 2006/01642, CC6) and two others in its C-terminal part (K220T, strains CLIP 1994/60344, CLIP 2000/80770 and CLIP 2001/87255, all ST13; and L221F, strain CLIP 1994/56373, CC1). These regions are known to be important for the binding of PrfA to PrfA-binding sites of target DNAs (25, 26). In addition, the A129P substitution, located between the β-roll and the hinge αD regions, occurred in a CC224 strain (CLIP 1994/58618). Finally, six of the PrfA$^-$ strains, all belonging to CC155, showed a reversion of the *prfA* stop codon due to the insertion of 5 nucleotides at position 712 in the *prfA* sequence, leading to a longer PrfA protein (238 amino acids in EGDe *vs* 293 amino acids in the CC155 strains of this study).

One of the four non-hemolytic mutants (CC1 strain CLIP 1998/76801) exhibited a wild-type (WT) PrfA sequence as compared to EGDe, but showed a PrfA$^-$ phenotype. This observation suggested that a mechanism interfering upstream of PrfA function was affected. Glutathione, synthetized by *Lm* through the glutathione synthase encoded by *gshF* (*lmo2770*), is critical for PrfA activation (28). Interestingly, although it is part of the *Lm* core genome (14, 21), *gshF* was absent in the genome of the CLIP 1998/76801 strain (**Fig. 1**) (see below), which could explain the absence of PrfA activity in this strain.

156

157 **Analysis of spontaneous LLO mutants**

158 Analysis of *hly* sequences in the 60 non-hemolytic strains identified multiple mutations
159 leading to amino acid substitutions in LLO (**Table S1**). Several substitutions (N31H, S35L,
160 V438I and K523S) were identified in at least 48 hemolytic *Lm* strains of our database
161 (~ 4,100 genomes), suggesting that they do not cause LLO loss of function. However, a
162 S250N substitution was only found in three non-hemolytic strains of this study (CLIP
163 2008/01432, 2008/01433 and 2008/01435, all CC77) and could therefore result in LLO loss
164 of function. Since these strains also expressed a truncated and non-active PrfA, which is
165 sufficient to explain the non-hemolytic phenotype of these strains, we did not pursue this
166 further.

167 Two out of the three non-hemolytic strains showing a WT PrfA sequence and a $PrfA^+$
168 phenotype (CC121 strains CLIP 2007/01406 and CLIP 2007/01014) exhibited a single amino
169 acid substitution in LLO (hly_{G299V} or LLO_{G299V}), which was not present in any of the other
170 strains. The third strain (CC2, CLIP 1989/13656) harbored a premature stop codon at position
171 484 in LLO (hly_{C484*} or LLO_{C484*}). The absence of any other specific feature in these three
172 strains that could be linked to the loss of hemolytic activity suggested that the G299V
173 mutation and the truncation at position 484 in LLO could be the cause of the loss of hemolytic
174 activity in these strains. In addition, two CC7 strains expressing a truncated PrfA (CLIP
175 1998/75799 and CLIP 1989/14490) also showed a premature stop codon in LLO at position
176 261 (hly_{N261*}) due to the insertion of one nucleotide.

177 In the CLIP 1998/76801 strain mentioned above, *hly* could not be detected by PCR and the
178 *hly* region could not be assembled from Illumina reads. In order to resolve this region, we
179 sequenced this strain using the single molecule, real-time (SMRT) sequencing technology
180 (Pacific Biosciences, California, USA). The CLIP 1998/76801 complete genome (CC1,

2.84 Mb) was compared to the closely related F2365 complete genome (CC1, NCBI accession number NC_002973) as reference. This showed that the LIPI-1 region had undergone an inversion of more than 40 kb (**Fig. 2A**). This large rearrangement splitted LIPI-1 into two parts with concomitant loss of *hly* and partial truncation of the 5' region of the adjacent *mpl* gene. Six ORFs were inserted upstream of *mpl* in CLIP 1998/76801 as compared to F2365, comprising genes encoding a transposition protein (*tnsB*) and a DNA invertase (*hin*), which are likely the cause of the rearrangement, as well as cadmium resistance genes (*cadA* and *cadC*) (**Fig. 2A**).

We confirmed that *gshF* is absent in CLIP 1998/76801, together with 12 other upstream and downstream genes related to sugar metabolism (**Fig. 2B**). These genes were replaced by 11 ORFs encoding a transposition protein (*tnsB*), a DNA invertase (*hin*) and cadmium resistance genes (*cadA* and *cadC*) similar to those inserted in the LIPI-1 region. In total, eight similar copies of this transposable element were found in the CLIP 1998/76801 genome, as well as many other large rearrangements and deletions (**Fig. 2C**). Similar transposable elements were detected in one *L. ivanovii* strain in NCBI database (accession number KR780025.1; 99% nucleotide identity, full length) and in 128 *Lm* strains (> 99.87% nucleotide similarity, full length) out of the 4,091 genome sequences available at the NRCL at the time of the study. These strains comprised 14.1% of all the CC1 strains (90/638, representing two distinct monophyletic groups within the phylogeny of CC1, data not shown) and all the CC59 strains (n = 38). No significant link of this element with food or clinical origins was found within CC1.

Assessment of *hly* and *prfA* transcription

In order to test the effect of the identified mutations on *hly* and *prfA* transcription, qRT-PCRs were performed for a representative set of non-hemolytic strains (one strain per type of loss-

of-hemolysis mutation, **Table S1**). All non-hemolytic strains showed *prfA* transcription levels equivalent to or higher than EGDe, except for strains CLIP 1998/75799 ($PrfA_{I51*}$-LLO_{N261*} mutations) and CLIP 1998/77604 ($PrfA_{T76*}$ mutation), which showed no amplification, likely due to poor primer annealing (8 mismatches with the *prfA*-R primer) (**Fig. S3**). As expected, strains with altered PrfA (aa substitution or truncation) showed no or extremely reduced *hly* transcription levels. These results show that for these strains the loss of hemolytic activity is due to *prfA* post-transcriptional events leading to the absence of PrfA activity. In the strain CLIP 2007/01406 (LLO_{G299V}), *hly* was transcribed at a similar level than in EGDe, whereas in CLIP 1989/13656 (LLO_{C484*}), *hly* transcription was weaker.

In vitro characterization of the hly_{G299V} and hly_{C484*} mutations

In order to characterize the functional impact of the G299V substitution (CLIP 2007/01406 and CLIP 2007/01014) and of the truncation at position 484 in LLO (CLIP 1989/13656), we introduced a plasmid containing either a wild type *hly* gene (hly_{WT}) or a mutated version of this gene (hly_{G299V} or hly_{C484*}, encoding LLO_{G299V} and LLO_{C484*}, respectively) in a EGDΔ*hly* strain. While EGDΔ*hly*:pPL2-hly_{WT} was hemolytic, EGDΔ*hly*:pPL2-hly_{G299V} or EGDΔ*hly*:pPL2-hly_{C484*} remained non-hemolytic, as assessed on Columbia horse blood agar-plates. These results demonstrate that the hly_{G299V} and hly_{C484*} mutations are responsible for the absence of hemolytic activity in the strains CLIP 2007/01406, CLIP 2007/01014 and CLIP 1989/13656.

Western blot analyses of culture supernatants detected lower amounts of LLO produced by EGDΔ*hly*:pPL2-hly_{G299V} and EGDΔ*hly*:pPL2-hly_{C484*} bacteria as compared to the WT EGD and EGDΔ*hly*:pPL2-hly_{WT} strains (**Fig. 3A**). qRT-PCR analyses showed that *hly* transcription level in both EGDΔ*hly*:pPL2-hly_{G299V} and EGDΔ*hly*:pPL2-hly_{C484*} strains is comparable to that observed in EGDΔ*hly*:pPL2-hly_{WT}, although slightly weaker for EGDΔ*hly*:pPL2-hly_{C484*}

(**Fig. 3B**). Furthermore, the EGDΔ*hly*:pPL2-hly_{C484*} mutant produced a shorter LLO protein as compared to strains harboring the hly_{WT}, confirming that the premature stop codon identified in *hly* in the CLIP 1989/13656 strain leads to the production of a truncated LLO. The hly_{N261*} mutation (**Fig. 1**; **Table S1**) was not tested *in vitro* as this premature stop codon is upstream of the hly_{C484*} mutation, leading to an even shorter LLO.

Virulence of hly_{G299V} and hly_{C484*} mutants

We finally assessed the virulence of the EGDΔ*hly*:pPL2-hly_{G299V} and EGDΔ*hly*:pPL2-hly_{C484*} complemented strains relative to that of the EGDΔ*hly*:pPL2-hly_{WT} and EGDΔ*hly*:pPL2 strains upon intravenous injection in mice. EGDΔ*hly*:pPL2-hly_{G299V} and EGDΔ*hly*:pPL2-hly_{C484*} strains were four order of magnitude less abundant than the EGDΔ*hly*:pPL2-hly_{WT} strain in the liver and the spleen (**Fig. 3C**). This demonstrates that the virulence of *Lm* expressing either LLO_{G299V} or LLO_{C484*} is strongly attenuated *in vivo*.

Discussion

Virulence gene polymorphisms leading to *Lm* attenuation have been previously described and have been associated with strains of lower pathogenic potential. The best characterized are those affecting the invasion-associated *inlA* gene, found in a large proportion (>25-30%) of lineage II food isolates but extremely rare among lineage I strains, more frequently associated with clinical cases (13, 21, 29, 30). Mutations leading to more radical *Lm* virulence attenuation have also been characterized, particularly those affecting the *prfA* gene (31-33), but their frequency and distribution across the *Lm* population remained undetermined. Here we examined the occurrence of "loss-of-virulence" mutations in *Lm* by screening a wide and diverse panel of strains for hemolytic activity. Our data shows that non-hemolytic *Lm* mutants occur at low frequency (0.1%) and are phylogenetically diverse, including strains belonging

to hypovirulent and hypervirulent clonal complexes (14). This indicates that the underlying mutational events are not linked to the genetic background of the strains.

Lm hemolytic phenotype depends on two essential virulence determinants, the central virulence regulator PrfA and LLO, encoded by *prfA* and *hly*, respectively. Indeed, all non-hemolytic strains identified in this study carried mutations in either of these genes. The large majority of non-hemolytic strains (56/57,820; 95%) carried *prfA* mutations (frameshifts, missense or nonsense nucleotide changes, or reversion of the stop codon into a glutamine codon). Although no PrfA activity could be detected and *hly* was not transcribed in these strains, *prfA* was transcribed at similar levels to EGDe. This suggests that the loss of PrfA activity in these strains likely results from PrfA misfolding, instability and/or inactivating amino-acid substitution. Some inactivating amino acid substitutions in PrfA occurred in the β-roll, HTH motif or C-terminal domain, in line with the critical role of these regions in PrfA activity (25-27, 31). As PrfA is the major transcriptional regulator of *Lm* virulence genes and is essential for its pathogenicity (23, 34), the virulence of PrfA$^-$ strains are expected to be highly attenuated as previously described (31-33). The first *Lm* strain naturally producing a C-terminally-extended PrfA polypeptide (55 residues longer) was identified in this study and showed no PrfA activity and no *hly* transcription.

Comparatively to PrfA, non-hemolytic *hly* mutants with affected LLO activity were less frequent (5/57,820; 0.01%) in our study. Our analysis identified for the first time a spontaneous amino-acid substitution in LLO (hly_{G299V}) and premature stop codons in *hly* (hly_{N261*} and hly_{C484*}) leading to the loss of LLO activity. Lower quantities of LLO were detected in the culture supernatants of the EGDΔ*hly*:pPL2-hly_{G299V} and EGDΔ*hly*:pPL2-hly_{C484*} constructs than for the EGD and EGDΔ*hly*:pPL2-hly_{WT} strains. The quantity of *hly* transcripts was similar in the EGDΔ*hly*:pPL2-hly_{G299V} and in the EGDΔ*hly*:pPL2-hly_{WT} control, indicating that LLO_{G299V} is likely less stable than WT LLO. In

281 contrast, EGDΔ*hly*:pPL2-hly_{C484*} showed a lower *hly* transcription level, suggesting an
282 impaired stability of hly_{C484*} transcript, relative to that of WT *hly*. *In vivo* experiments
283 confirmed that the non-hemolytic strains harboring the hly_{G299V} or hly_{C484*} mutations have
284 strongly attenuated virulence in mice. In line with these results, only three non-hemolytic
285 strains were isolated from human samples. Although we did not have access to the detailed
286 clinical data of these patients (dating back from the 1980s and 90s), one possibility would be
287 that they were heavily immunocompromised, mirroring previous reports on isolation of the
288 non-pathogenic *Lm* relative *Listeria innocua* from immunosuppressed individuals (35).
289 One of the LLO^- strains (CLIP 1998/76801) underwent huge genomic rearrangements that
290 likely caused the loss of *hly* and *gshF,* encoding a glutathione synthase reported as being
291 required for PrfA activity (28). CLIP 1998/76801 is the only strain in our entire genome
292 database (~ 4,100 entries) that lacks *gshF*. Interestingly, each copy of the transposable
293 element that likely caused the genomic rearrangements observed in this strain carried putative
294 cadmium resistance determinants that could be advantageous in environments in which
295 virulence determinants are not needed. Similar transposable elements were detected in
296 monophyletic groups of CC1 and CC59 strains, suggesting that it has been horizontally
297 transmitted in the *Lm* population.
298 The predominance of $PrfA^-$ mutants among the non-hemolytic strains could reflect the fact
299 that *prfA* is a pleiotropic regulatory gene that controls the expression of a number of virulence
300 determinants, the expression of which is known to entail a significant fitness cost in non-host
301 conditions (24). Our results show that, at 22°C, the majority of $PrfA^-$ strains have similar
302 fitness than EGDe, suggesting that the absence of PrfA activity does not impact *Lm* fitness in
303 non-pathogenic conditions. Nevertheless, a reduced fitness was observed at 37°C (mammalian
304 host temperature), comparatively to EGDeΔ*prfA*. This result suggests that non-hemolytic
305 strains are more adapted to a non-pathogenic lifestyle, independently of PrfA. Consistent with

306 this, most of the non-hemolytic *Lm* isolates were from non-clinical origins. The ratio of non-
307 synonymous and synonymous substitutions (dN/dS) estimated for *prfA* (dN/dS=0.08892) and
308 *hly* (dN/dS=0.03674) using a dataset of 100 genomes representative of *Lm* phylogenetic
309 diversity (14) confirmed that, similarly to *Lm* core genes (dN/dS=0.05353 in average, (21)),
310 these genes are under purifying selection. Thus, any deleterious mutations affecting these
311 genes tend to be eliminated from *Lm* population. The exceptionally low frequency of
312 deleterious mutations in *prfA* and *hly* indicates that there is a strong necessity for *Lm* to retain
313 its virulence capacity (36). Our results also suggest that, although exceptionally, once strains
314 loss their virulence capacity (*e.g.* due to a *prfA* mutation), other virulence genes may become
315 unneeded and prone to accumulate mutations, as observed in our $PrfA^-/LLO^-$ and $PrfA^-/GshF^-$
316 strains. Previous studies have already identified strains with multiple mutations occurring in
317 several major virulence genes (20). Strains with virulence attenuating mutations are therefore
318 prone to enter into an evolutionary path towards obligate saprophytism. The *Lm*
319 phylogenomic clade comprises another pathogenic species, *Listeria ivanovii*, with a full
320 complement of PrfA-regulated genes, as well as non-pathogenic species, some of which
321 contain remnants thereof (e.g. *Listeria seeligeri* or *L. innocua*) (37, 38). While infrequent,
322 spontaneous virulence-disabling mutations, as those described here, could have been key
323 initial events in the emergence and evolution of the *Lm*-related non-pathogenic *Listeria*
324 species.

325

326 **Materials and methods**

327 **Bacterial strains and growth media**

328 The 60 non-hemolytic *Lm* isolates included in this study were identified among a collection of
329 57,820 *Lm* strains collected between 1987 and 2008 by the French National Reference Centre
330 for *Listeria* (NRCL) and World Health Organization Collaborating Centre for *Listeria*

(WHOCCL) in the context of the epidemiological surveillance of listeriosis. This global collection included isolates of food (n = 36,630), clinical (n = 5,980), environmental (n = 3,647), veterinary (n = 1,713) and unknown (n = 9,850) origins. Isolates were revived by plating them onto Columbia Agar and single colonies were grown on Columbia Agar slants. *Lm* strains were routinely grown in BHI at 37°C and *Escherichia coli* strains were grown at 37°C in LB broth or agar plates.

Phenotypic characterization of *Listeria* isolates

Miniaturized enzymatic and sugar fermentation tests (API-*Listeria* identification microgallery, BioMérieux, France), in combination with the hemolytic activity assessment of strains, were used for phenotypic identification of *Listeria* species (39). Hemolytic activity was tested on Columbia horse blood agar-plates (BioMérieux, France). *Lm* CLIP 74910 and *Listeria innocua* CLIP 74915 were used as positive and negative controls of hemolysis, respectively.

Genome sequencing and analyses

Genomic DNA was extracted using the DNeasy Blood and Tissue extraction kit (Qiagen, Denmark) and used for whole genome sequencing on Illumina NextSeq 500 (2 x 150 bp) platform (Illumina, California, USA). Reads were trimmed with AlienTrimmer (Criscuolo & Brisse 2013) to eliminate adapter sequences and discard reads with Phred scores of ≤ 20. *De novo* assembly of Illumina reads was performed using SPAdes Genome Assembler 3.1 (40). The complete genome of the CLIP 1998/76801 strain was obtained by PacBio RS II sequencing (Pacific Biosciences, California, USA) using DNA purified with the Wizard genomic DNA purification kit (Promega, Wisconsin, USA). Genome annotation was performed using Prokka 1.11 (41).

PCR-serogroups (12, 42), MLST profiles (13) and cgMLST profiles (21) were deduced from genome assemblies using the BIGSdb-*Lm* platform (http://bigsdb.pasteur.fr/listeria; (21)). Genome assemblies were made publicly available in the BIGSdb-*Lm* platform (**Table S1**).

Assessment of *prfA* and *hly* evolutionary trends

prfA and *hly* sequences were extracted from 100 genomes that were selected to represent the species diversity based on MLST and PFGE typing (14) and aligned using Muscle 3.8 (43). This dataset included genomes from 13 food isolates, 45 human clinical isolates, 19 animal isolates, 1 environmental isolate and 22 isolates of unknown origin. They comprised 41 genomes of lineage I, 53 of lineage II, 5 of lineage III, and 1 of lineage IV and represented 5 singletons and 34 clonal complexes based on MLST. No non-hemolytic strain was included in this analysis. Alignments were used to estimate the non-synonymous and synonymous ratios (dN/dS) of *prfA* and *hly* using the *codeml* program, included in the PAML 4.4 package (44).

Assessment of PrfA activity

PrfA activity was assessed by measuring the activity of PrfA-regulated *plcB* and *hpt* gene products as previously described (45, 46). For PlcB, lecithinase tests were performed in egg-yolk BHI, for Hpt, glucose-1-phosphate acidification tests were carried out in phenol red broth, in both cases with and without 0.5% w/v activated charcoal (Merck, New Jersey, USA). Medium supplementation with charcoal leads to the partial activation of PrfA, presumably due to sequestration of repressor substances from the culture medium. Three *Lm* genotypes from strain P14 were used as controls: (i) $prfA_{WT}$ characterized by an activable PrfA phenotype (lack of PlcB and Hpt activity in normal medium and strong activity in charcoal-supplemented medium), (ii) Δ*prfA* which remains negative for PlcB and Hpt activity

in the presence of charcoal; and (iii) constitutively activated *prfA** with strong PlcB and Hpt activity independently of charcoal supplementation (24, 45, 47).

RNA extractions

Non-hemolytic strains and EGDΔ*hly*:pPL2-hly_{WT}, EGDΔ*hly*:pPL2, EGDΔ*hly*:pPL2-hly_{G299V} and EGDΔ*hly*:pPL2-hly_{C484*} constructs were cultured overnight on BHI agar at 37°C. One colony was used to inoculate 5 ml of BHI broth. After overnight growth at 37°C, 500µl of culture was added to 10ml of BHI broth and the whole exponential phase culture (at 37°C) was centrifuged at 5,000 g for 5 min. The pellet was suspended with 400 µl of resuspension buffer (10% glucose, 12.5 mM TRIS, 10 mM EDTA in nuclease-free water) and transferred to a lysing tube (containing 0.1 mm of ceramic breads, 500 µl of acid phenol and 60 µl of EDTA 0.5 M). The Precellys24 homogenizer (Bertin Instruments, France) was used at 6,500 rpm, for 2 x 23 s (10 s break), and the resulting mixture was centrifuged at 14,000 g at 4°C for 10 min. The upper aqueous phase was transferred into a tube containing 1 ml of Trizol and 100 µl of chloroform, mixed by inversions and centrifuged. The upper aqueous phase was transferred into a tube containing 200 µl of chloroform, mixed by inversions and centrifuged. The upper aqueous phase was transferred into a storage tube (containing 650 µl of isopropanol and 65 µl sodium acetate 3 M), mixed by inversions, precipitated 20 min at -20°C and centrifuged during 20 min. The supernatant was rinsed twice with ethanol 75%. The air-dried pellet was dissolved in 300 µl of nuclease-free water. RNA concentrations were measured with the DeNovix DS-11 Spectrophotometer (DeNovix, Delaware, USA) and diluted to obtain 500 ng of RNA in 12.5 µl of nuclease-free water.

Quantification of *hly* and *prfA* transcripts by qRT-PCR

For the qRT-PCRs, cDNAs were generated prior to qPCRs. DNase treatment was performed with the RNase-free DNase I (New England BioLabs, Massachusetts, USA) according to the instructions of the manufacturer (by adding 0.5 µl of RNAseOUT, 0.5 µl of DNAseI and 1.5 µl of buffer 10x to the 12.5 µl of diluted RNAs; then 1.5 µl of EDTA 0.05 M). cDNAs were generated by reverse transcription using the M-MLV Reverse Transcriptase (Invitrogen, California, USA) and random hexamers for priming according to the instructions of the manufacturer (by adding 2 µl of dNTPs 10 mM, 2 µl of random primers 2.5 µM and 3.5 µl of nuclease-free water to the 16.5 µl of previously DNase treated sample; then 8 µl of First-Strand buffer 5x, 4 µl of DTT 0.1 M and 2 µl of nuclease-free water; then 0.5 µl of M-MLV RT).

All quantitative PCRs were prepared using SYBR® Green Real-Time PCR Master Mixes and StepOnePlus™ Real-Time PCR System (Applied Biosystemss, California, USA). Each primer pair was used in separated reactions using PCR mixtures containing 1 µl of each primer 9 µM (**Table S2**), 5 µl of Sybr mix, 1 µl of cDNA diluted at 1:5 and 3 µl of nuclease-free water. Real-time PCR reactions were carried out in MicroAmp™ Fast Optical 96-Well Reaction Plates (Applied Biosystems, California, USA) using the following protocol: initial denaturation at 95°C for 10 min, then 40 cycles of denaturation at 95°C for 15 s and primer annealing/elongation at 60°C for 1 min. Each strain was tested at least three times using independent pre-cultures. *gyrB* was used as stable reference gene for normalization. Results are shown as fold change of target gene expression relative to EGDe or EGD (Relative Quantities, RQs), which were deduced from the cycle threshold values (CTs) using the $2^{-\Delta\Delta CT}$ methodology.

Fitness studies

The microbial growth of non-hemolytic strains, EGDe, EGD and EGDeΔ*prfA* was monitored over time in BHI at 22°C and 37°C using absorbance measurements (OD_{600nm}) through the Bioscreen C® system (Oy Growth Curves Ab Ltd, Helsinki, Finland). Bacteria were first cultured overnight on BHI agar at 22°C or 37°C and one colony was used to inoculate 5 ml of BHI broth. After overnight growth, the stationary phase cultures were diluted to reach an OD_{600nm} of 0.1 and transferred into Bioscreen C® 96-well plates. OD_{600nm} of non-inoculated wells (blanks) were subtracted from inoculated ones to delete the background noise. Each strain was tested three times. Mean OD_{600nm} per strain were used to calculate area under the curves over time. For this, data were fitted to parametric models (Gompertz, modified Gompertz, Logistic and Richards laws) using the "gcFit" function of the "grofit" R package v.1.1.1-1 (48). The model that best fitted the data was selected by means of an AIC criterion (49) and used to derive areas under the growth curves.

DNA manipulations and cloning

We used a two-step cloning strategy to introduce the wild type *hly* (hly_{WT}), hly_{G299V} or hly_{C484*} in the *Lm* strain EGDΔ*hly*. First, we cloned separately the hly_{WT}, hly_{G299V} and hly_{C484*} gene sequences into the *Listeria* integrative vector pPL2 (50). Primers used are listed in the **Table S2**. To deliver plasmids into *Lm*, *Escherichia coli* S17.1 (colistin and nalidixic acid sensitive) were transformed with the plasmids followed by conjugation with *Lm* EGDΔ*hly* (colistin and nalidixic acid resistant). *Lm* EGDΔ*hly* were selected on 7µg/ml chloramphenicol (bacteria containing the pPL2 derivatives), 10µg/ml colicin and 50µg/ml nalidixic acid (selection of resistant *Lm vs* sensitive *E. coli*). Since all our constructs were made on a similar EGD background, no impact of the PrfA* phenotype of EGD is expected on our results and conclusions.

453 **Western blotting**

454 Protein extracts were obtained from EGD, EGDΔ*hly*, EGDΔ*hly*:pPL2, EGDΔ*hly*:pPL2-hly_{WT},
455 EGDΔ*hly*:pPL2-hly_{G299V} and EGDΔ*hly*:pPL2-hly_{C484*} as follows. Bacteria were grown
456 overnight in BHI broth at 37°C. After centrifugation of bacterial cultures (30 min; 2,151 g),
457 all proteins of the supernatant were precipitated by using trichloroacetic acid (20%) and
458 washed using acetone. Proteins were then separated by SDS/PAGE (8% acrylamide gel and
459 3.9% stacking gel) and transferred to a polyvinylidene difluoride transfer membrane (Bio-
460 Rad, California, USA). The membrane was incubated overnight at 4°C with a blocking buffer
461 containing dried milk (5%), phosphate buffered saline (PBS, 1%) and Tween (0.1%) and
462 washed with PBS (1%) and Tween (0.1%). It was then incubated first with a polyclonal anti-
463 LLO (51, 52) or anti-InlC antibody (53) (1/20,000; 1h; room temperature) and second with
464 the anti-rabbit antibody (1/3,000; 1h; room temperature). The membrane was washed with
465 PBS (1%) and Tween (0.1%) between each incubation step with antibodies. Antibody-antigen
466 interactions were revealed using a SuperSignal West Pico Chemiluminescent substrate
467 (Thermo Fischer Scientific, Massachussetts, USA).

468

469 **Animal studies**

470 The virulence of *Lm* strains EGDΔ*hly*:pPL2-hly_{WT}, EGDΔ*hly*:pPL2-hly_{G299V},
471 EGDΔ*hly*:pPL2-hly_{C484*} and EGDΔ*hly*:pPL2 was assessed *in vivo*. Balb/c mice were infected
472 via intravenous route with 1.10^4 colony-forming units (CFUs) per animal. At 72 h post
473 infection, mice were sacrificed for spleen and liver dissection. CFUs were enumerated by
474 plating dilutions of the whole homogenized organs onto BHI plates. Statistical analyses were
475 done using the Mann–Whitney *U* test as compared with EGDΔ*hly*:pPL2-hly_{WT}.

476

477 **Funding information**

Work in M. Lecuit's laboratory was supported by Institut Pasteur, Inserm, LabEx IBEID, the European Research Council, and Santé Publique France. Work in J. Vázquez-Boland's laboratory was supported by the Wellcome Trust (programme grant WT074020MA) and partially by core Roslin Institute funding from BBSRC (BB/J004227/1). The funders had no role in study design, data collection and interpretation, or the decision to submit the work for publication.

Acknowledgments

We acknowledge Anne Morvan, Thomas Cantinelli, Nathalie Tessaud-Rita and Laetitia Bellon for their help in strain collection and isolation at the National Reference Centre for *Listeria*. We thank Pascale Cossart for providing antibodies against LLO, and the Institut Pasteur P2M platform for genome sequencing.

490

491 **References**

492 1. Freitag NE, Port GC, Miner MD. 2009. *Listeria monocytogenes* - from saprophyte to
493 intracellular pathogen. Nat Rev Microbiol 7:623-8.

494 2. Cossart P. 2011. Illuminating the landscape of host-pathogen interactions with the
495 bacterium *Listeria monocytogenes*. Proc Natl Acad Sci U S A 108:19484-91.

496 3. Portnoy DA, Jacks PS, Hinrichs DJ. 1988. Role of hemolysin for the intracellular
497 growth of *Listeria monocytogenes*. J Exp Med 167:1459-71.

498 4. Hamon MA, Ribet D, Stavru F, Cossart P. 2012. Listeriolysin O: the Swiss army knife
499 of *Listeria*. Trends Microbiol 20:360-8.

500 5. Vazquez-Boland JA, Dominguez-Bernal G, Gonzalez-Zorn B, Kreft J, Goebel W.
501 2001. Pathogenicity islands and virulence evolution in *Listeria*. Microbes Infect
502 3:571-84.

503 6. Leimeister-Wachter M, Haffner C, Domann E, Goebel W, Chakraborty T. 1990.
504 Identification of a gene that positively regulates expression of listeriolysin, the major
505 virulence factor of *Listeria monocytogenes*. Proc Natl Acad Sci U S A 87:8336-40.

506 7. Chakraborty T, Leimeister-Wachter M, Domann E, Hartl M, Goebel W, Nichterlein T,
507 Notermans S. 1992. Coordinate regulation of virulence genes in *Listeria*
508 *monocytogenes* requires the product of the *prfA* gene. J Bacteriol 174:568-74.

509 8. Piffaretti JC, Kressebuch H, Aeschbacher M, Bille J, Bannerman E, Musser JM,
510 Selander RK, Rocourt J. 1989. Genetic characterization of clones of the bacterium
511 *Listeria monocytogenes* causing epidemic disease. Proc Natl Acad Sci U S A 86:3818-
512 22.

513 9. Wiedmann M, Bruce JL, Keating C, Johnson AE, McDonough PL, Batt CA. 1997.
514 Ribotypes and virulence gene polymorphisms suggest three distinct *Listeria*

515 *monocytogenes* lineages with differences in pathogenic potential. Infect Immun
516 65:2707-16.
517 10. Orsi RH, den Bakker HC, Wiedmann M. 2011. *Listeria monocytogenes* lineages:
518 Genomics, evolution, ecology, and phenotypic characteristics. Int J Med Microbiol
519 301:79-96.
520 11. Seeliger H.P.R. JD. 1986. in Bergey's Manual of Systematic Bacteriology (Williams
521 & Wilkins). 2:1235–1245.
522 12. Doumith M, Buchrieser C, Glaser P, Jacquet C, Martin P. 2004. Differentiation of the
523 major *Listeria monocytogenes* serovars by multiplex PCR. J Clin Microbiol 42:3819-
524 22.
525 13. Ragon M, Wirth T, Hollandt F, Lavenir R, Lecuit M, Le Monnier A, Brisse S. 2008. A
526 new perspective on *Listeria monocytogenes* evolution. PLoS Pathog 4:e1000146.
527 14. Maury MM, Tsai YH, Charlier C, Touchon M, Chenal-Francisque V, Leclercq A,
528 Criscuolo A, Gaultier C, Roussel S, Brisabois A, Disson O, Rocha EP, Brisse S,
529 Lecuit M. 2016. Uncovering *Listeria monocytogenes* hypervirulence by harnessing its
530 biodiversity. Nat Genet 48:308-13.
531 15. Tabouret M, De Rycke J, Audurier A, Poutrel B. 1991. Pathogenicity of *Listeria*
532 *monocytogenes* isolates in immunocompromised mice in relation to listeriolysin
533 production. J Med Microbiol 34:13-8.
534 16. Allerberger F, Dierich M, Petranyi G, Lalic M, Bubert A. 1997. Nonhemolytic strains
535 of *Listeria monocytogenes* detected in milk products using VIDAS immunoassay kit.
536 Zentralbl Hyg Umweltmed 200:189-95.
537 17. Moreno LZ, Paixao R, de Gobbi DD, Raimundo DC, Porfida Ferreira TS, Micke
538 Moreno A, Hofer E, dos Reis CM, Matte GR, Matte MH. 2014. Phenotypic and

539 genotypic characterization of atypical *Listeria monocytogenes* and *Listeria innocua*
540 isolated from swine slaughterhouses and meat markets. Biomed Res Int 2014:742032.
541 18. Palerme JS, Pan PC, Parsons CT, Kathariou S, Ward TJ, Jacob ME. 2016. Isolation
542 and characterization of atypical *Listeria monocytogenes* associated with a canine
543 urinary tract infection. J Vet Diagn Invest 28:604-7.
544 19. Burall LS, Grim C, Gopinath G, Laksanalamai P, Datta AR. 2014. Whole-Genome
545 Sequencing Identifies an Atypical *Listeria monocytogenes* Strain Isolated from Pet
546 Foods. Genome Announc 2.
547 20. Roche SM, Grepinet O, Kerouanton A, Ragon M, Leclercq A, Temoin S, Schaeffer B,
548 Skorski G, Mereghetti L, Le Monnier A, Velge P. 2012. Polyphasic characterization
549 and genetic relatedness of low-virulence and virulent *Listeria monocytogenes* isolates.
550 BMC Microbiol 12:304.
551 21. Moura A, Criscuolo A, Pouseele H, Maury MM, Leclercq A, Tarr C, Bjorkman JT,
552 Dallman T, Reimer A, Enouf V, Larsonneur E, Carleton H, Bracq-Dieye H, Katz LS,
553 Jones L, Touchon M, Tourdjman M, Walker M, Stroika S, Cantinelli T, Chenal-
554 Francisque V, Kucerova Z, Rocha EP, Nadon C, Grant K, Nielsen EM, Pot B, Gerner-
555 Smidt P, Lecuit M, Brisse S. 2016. Whole genome-based population biology and
556 epidemiological surveillance of *Listeria monocytogenes*. Nat Microbiol 2:16185.
557 22. Johansson J, Mandin P, Renzoni A, Chiaruttini C, Springer M, Cossart P. 2002. An
558 RNA thermosensor controls expression of virulence genes in *Listeria monocytogenes*.
559 Cell 110:551-61.
560 23. Scortti M, Monzo HJ, Lacharme-Lora L, Lewis DA, Vazquez-Boland JA. 2007. The
561 PrfA virulence regulon. Microbes Infect 9:1196-207.

562 24. Vasanthakrishnan RB, de Las Heras A, Scortti M, Deshayes C, Colegrave N,
563 Vazquez-Boland JA. 2015. PrfA regulation offsets the cost of *Listeria* virulence
564 outside the host. Environ Microbiol 17:4566-79.
565 25. Vega Y, Rauch M, Banfield MJ, Ermolaeva S, Scortti M, Goebel W, Vazquez-Boland
566 JA. 2004. New *Listeria monocytogenes prfA** mutants, transcriptional properties of
567 PrfA* proteins and structure-function of the virulence regulator PrfA. Mol Microbiol
568 52:1553-65.
569 26. Herler M, Bubert A, Goetz M, Vega Y, Vazquez-Boland JA, Goebel W. 2001.
570 Positive selection of mutations leading to loss or reduction of transcriptional activity
571 of PrfA, the central regulator of *Listeria monocytogenes* virulence. J Bacteriol
572 183:5562-70.
573 27. Deshayes C, Bielecka MK, Cain RJ, Scortti M, de las Heras A, Pietras Z, Luisi BF,
574 Nunez Miguel R, Vazquez-Boland JA. 2012. Allosteric mutants show that PrfA
575 activation is dispensable for vacuole escape but required for efficient spread and
576 *Listeria* survival *in vivo*. Mol Microbiol 85:461-77.
577 28. Reniere ML, Whiteley AT, Hamilton KL, John SM, Lauer P, Brennan RG, Portnoy
578 DA. 2015. Glutathione activates virulence gene expression of an intracellular
579 pathogen. Nature 517:170-3.
580 29. Jacquet C, Doumith M, Gordon JI, Martin PM, Cossart P, Lecuit M. 2004. A
581 molecular marker for evaluating the pathogenic potential of foodborne *Listeria*
582 *monocytogenes*. J Infect Dis 189:2094-100.
583 30. Nightingale KK, Windham K, Martin KE, Yeung M, Wiedmann M. 2005. Select
584 *Listeria monocytogenes* subtypes commonly found in foods carry distinct nonsense
585 mutations in *inlA*, leading to expression of truncated and secreted internalin A, and are

associated with a reduced invasion phenotype for human intestinal epithelial cells. Appl Environ Microbiol 71:8764-72.

31. Velge P, Herler M, Johansson J, Roche SM, Temoin S, Fedorov AA, Gracieux P, Almo SC, Goebel W, Cossart P. 2007. A naturally occurring mutation K220T in the pleiotropic activator PrfA of *Listeria monocytogenes* results in a loss of virulence due to decreasing DNA-binding affinity. Microbiology 153:995-1005.

32. Miner MD, Port GC, Bouwer HG, Chang JC, Freitag NE. 2008. A novel *prfA* mutation that promotes *Listeria monocytogenes* cytosol entry but reduces bacterial spread and cytotoxicity. Microb Pathog 45:273-81.

33. Rupp S, Aguilar-Bultet L, Jagannathan V, Guldimann C, Drogemuller C, Pfarrer C, Vidondo B, Seuberlich T, Frey J, Oevermann A. 2015. A naturally occurring *prfA* truncation in a *Listeria monocytogenes* field strain contributes to reduced replication and cell-to-cell spread. Vet Microbiol 179:91-101.

34. de las Heras A, Cain RJ, Bielecka MK, Vazquez-Boland JA. 2011. Regulation of *Listeria* virulence: PrfA master and commander. Curr Opin Microbiol 14:118-27.

35. Perrin M, Bemer M, Delamare C. 2003. Fatal case of *Listeria innocua* bacteremia. J Clin Microbiol 41:5308-9.

36. Bruno JC, Jr., Freitag NE. 2011. *Listeria monocytogenes* adapts to long-term stationary phase survival without compromising bacterial virulence. FEMS Microbiol Lett 323:171-9.

37. den Bakker HC, Cummings CA, Ferreira V, Vatta P, Orsi RH, Degoricija L, Barker M, Petrauskene O, Furtado MR, Wiedmann M. 2010. Comparative genomics of the bacterial genus *Listeria*: Genome evolution is characterized by limited gene acquisition and limited gene loss. BMC Genomics 11:688.

610 38. Orsi RH, Wiedmann M. 2016. Characteristics and distribution of *Listeria* spp.,
611 including *Listeria* species newly described since 2009. Appl Microbiol Biotechnol
612 100:5273-87.
613 39. Fujisawa T, Mori M. 1994. Evaluation of media for determining hemolytic activity
614 and that of API *Listeria* system for identifying strains of *Listeria monocytogenes*. J
615 Clin Microbiol 32:1127-9.
616 40. Bankevich A, Nurk S, Antipov D, Gurevich AA, Dvorkin M, Kulikov AS, Lesin VM,
617 Nikolenko SI, Pham S, Prjibelski AD, Pyshkin AV, Sirotkin AV, Vyahhi N, Tesler G,
618 Alekseyev MA, Pevzner PA. 2012. SPAdes: a new genome assembly algorithm and
619 its applications to single-cell sequencing. J Comput Biol 19:455-77.
620 41. Seemann T. 2014. Prokka: rapid prokaryotic genome annotation. Bioinformatics
621 30:2068-9.
622 42. Leclercq A, Chenal-Francisque V, Dieye H, Cantinelli T, Drali R, Brisse S, Lecuit M.
623 2011. Characterization of the novel *Listeria monocytogenes* PCR serogrouping profile
624 IVb-v1. Int J Food Microbiol 147:74-7.
625 43. Edgar RC. 2004. MUSCLE: multiple sequence alignment with high accuracy and high
626 throughput. Nucleic Acids Res 32:1792-7.
627 44. Yang Z. 2007. PAML 4: phylogenetic analysis by maximum likelihood. Mol Biol
628 Evol 24:1586-91.
629 45. Ermolaeva S, Karpova T, Novella S, Wagner M, Scortti M, Tartakovskii I, Vazquez-
630 Boland JA. 2003. A simple method for the differentiation of *Listeria monocytogenes*
631 based on induction of lecithinase activity by charcoal. Int J Food Microbiol 82:87-94.
632 46. Chico-Calero I, Suarez M, Gonzalez-Zorn B, Scortti M, Slaghuis J, Goebel W,
633 Vazquez-Boland JA, European *Listeria* Genome C. 2002. Hpt, a bacterial homolog of

the microsomal glucose-6-phosphate translocase, mediates rapid intracellular proliferation in *Listeria*. Proc Natl Acad Sci U S A 99:431-6.

47. Ermolaeva S, Novella S, Vega Y, Ripio MT, Scortti M, Vazquez-Boland JA. 2004. Negative control of *Listeria monocytogenes* virulence genes by a diffusible autorepressor. Mol Microbiol 52:601-11.

48. Kahm M, Hasenbrink G, Lichtenberg-Frate H, Ludwig J, Kschischo M. 2010. grofit: Fitting Biological Growth Curves with R. Journal of Statistical Software 33:1-21.

49. Akaike H. 1974. A new look at the statistical model identification. IEEE Transactions on Automatic Control 19:716-723.

50. Lauer P, Chow MY, Loessner MJ, Portnoy DA, Calendar R. 2002. Construction, characterization, and use of two *Listeria monocytogenes* site-specific phage integration vectors. J Bacteriol 184:4177-86.

51. Gouin E, Dehoux P, Mengaud J, Kocks C, Cossart P. 1995. *iactA* of *Listeria ivanovii*, although distantly related to *Listeria monocytogenes actA*, restores actin tail formation in an *L. monocytogenes actA* mutant. Infect Immun 63:2729-37.

52. Ribet D, Hamon M, Gouin E, Nahori MA, Impens F, Neyret-Kahn H, Gevaert K, Vandekerckhove J, Dejean A, Cossart P. 2010. *Listeria monocytogenes* impairs SUMOylation for efficient infection. Nature 464:1192-5.

53. Archambaud C, Gouin E, Pizarro-Cerda J, Cossart P, Dussurget O. 2005. Translation elongation factor EF-Tu is a target for Stp, a serine-threonine phosphatase involved in virulence of *Listeria monocytogenes*. Mol Microbiol 56:383-96.

54. Altschul SF, Madden TL, Schaffer AA, Zhang J, Zhang Z, Miller W, Lipman DJ. 1997. Gapped BLAST and PSI-BLAST: a new generation of protein database search programs. Nucleic Acids Res 25:3389-402.

658 55. Sullivan MJ, Petty NK, Beatson SA. 2011. Easyfig: a genome comparison visualizer.
659 Bioinformatics 27:1009-10.
660

Figure legends

Figure 1: Phylogenetic tree summarizing all the genetic features causing the loss of hemolytic activity among the 60 non-hemolytic *Lm* strains.

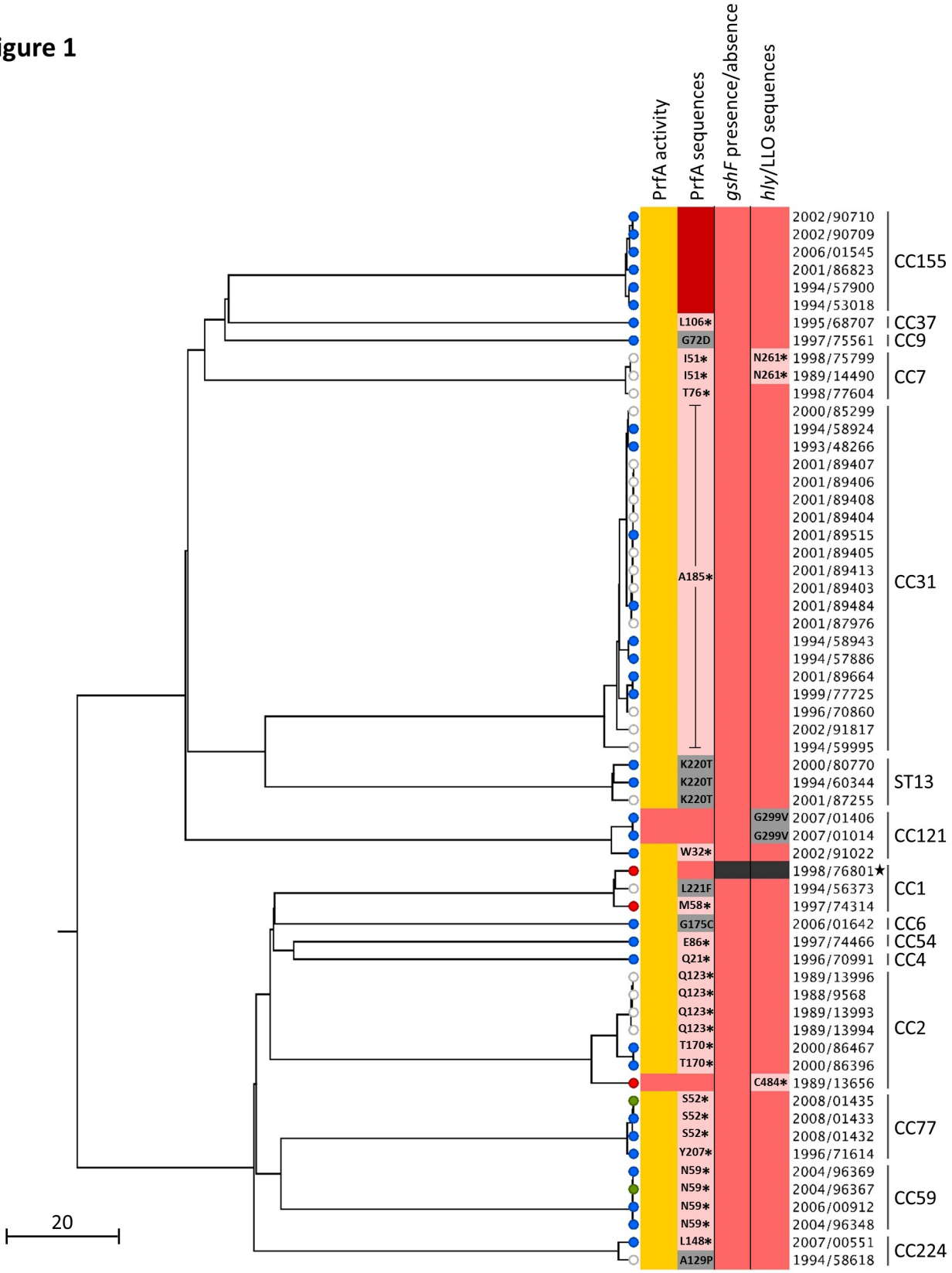
Single-linkage based clustering was obtained based on the cgMLST allelic profiles, as described previously (21). Scale bar indicates % of cgMLST similarity. Strain names have been simplified to avoid redundancy and should be preceded by "CLIP". PrfA activities and mutations (first and second columns, respectively), *gshF* presence/absence profile (third column) and LLO mutations and presence/absence profile (fourth column) are mapped on the phylogeny. Position and nature of amino acid substitutions are indicated in grey zones. Positions of premature stop codons are indicated next to black asterisks in light pink zones. Absence of *gshF* and *hly* in the CLIP 1998/76801 strain is indicated in black. MLST clonal complexes are shown on the right. The black star highlights the CLIP 1998/76801 strain that contains multiple copies of a transposable element that induced huge genomic rearrangements. ND: unknown and non-human origin.

Figure 2: Comparison of the CLIP 1998/76801 and F2365 genomes

A. Gene content of the LIPI-1 region in F2365 (accession number: NC_002973) (top) in comparison to the corresponding region in the non-hemolytic CLIP 1998/76801 strain (bottom). LIPI-1 genes are highlighted in red. *mpl* is composed of 1532 bp in F2365 against 1133 bp in CLIP 1998/76801. B. Gene content of the *gshF* region in F2365 (top) in comparison to the corresponding region in CLIP 1998/76801 (bottom). In A. and B., genes that are present in CLIP 1998/76801 but absent in F2365 are indicated in orange. Genes encoding the transposition protein (*tnsB*), the DNA-invertase (*hin*) and the cadmium resistance genes (*cadA* and *cadC*) are indicated. C. Global comparison of the F2365 (top) and

685 the CLIP 1998/76801 (bottom) genomes. Positions of the eight copies of the transposable
686 element are indicated in dark blue. Identity percentages (indicated by grey zones of variable
687 intensities) between sequences were determined by nucleotide BLAST (54). Genome
688 comparisons were performed using Easyfig 2.1 (55).
689
690 **Figure 3: Characterization of the G299V substitution in LLO and the truncated LLO at**
691 **positon 484.**
692 A. Western blotting of the culture supernatants of EGD and EGDΔ*hly* complemented or not
693 with the pPL2 plasmid alone or containing hly_{WT}, hly_{G299V} or hly_{C484*}. LLO detection was
694 performed by using LLO-specific antibodies (above) and InlC-specific antibodies were used
695 as loading controls (below). B. qRT-PCR quantification of *hly* transcripts produced in BHI at
696 37°C by the EGDΔ*hly* strain complemented with the pPL2 plasmid alone or containing the
697 hly_{WT}, hly_{G299V} or hly_{C484*} genes. Each strain was tested at least three times using independent
698 pre-cultures. *gyrB* was used as stable reference gene for normalization. Results are shown as
699 fold change of *hly* expression relative to EGD (Relative Quantities, RQs). Each central bar
700 represents the mean of at least three replications. Error bars indicate standard deviations from
701 the means. C. *In vivo* characterization of the hly_{G299V} and hly_{C484*} mutations as compared to
702 the hly_{WT}. Each Balb/C mice were infected intravenously with 1.10^4 CFUs. Animals were
703 sacrificed 72 h after infection. Numbers of CFUs per organ are shown for all the tested
704 strains. No bacteria could be recovered from the liver of mice infected by
705 EGDΔ*hly*:pPL2-hly_{G299V} and EGDΔ*hly*:pPL2 strains; and from the spleen of mice infected by
706 EGDΔ*hly*:pPL2. Statistical analyses were done by a Mann-Whitney *U* test as compared with
707 EGDΔ*hly*:pPL2-hly_{WT}.

Figure 1



Origin of isolates
- Food
- Human
- Production environment
- ND

PrfA activity
- PrfA+
- PrfA-

PrfA sequences
- No loss-of-function mutation
- Larger protein
- Truncated
- Amino acid substitution

***gshF* presence/absence**
- Present
- Missing gene

***hly*/LLO sequences**
- No loss-of-function mutation
- Truncated protein
- Amino acid substitution
- Missing gene

Figure 2

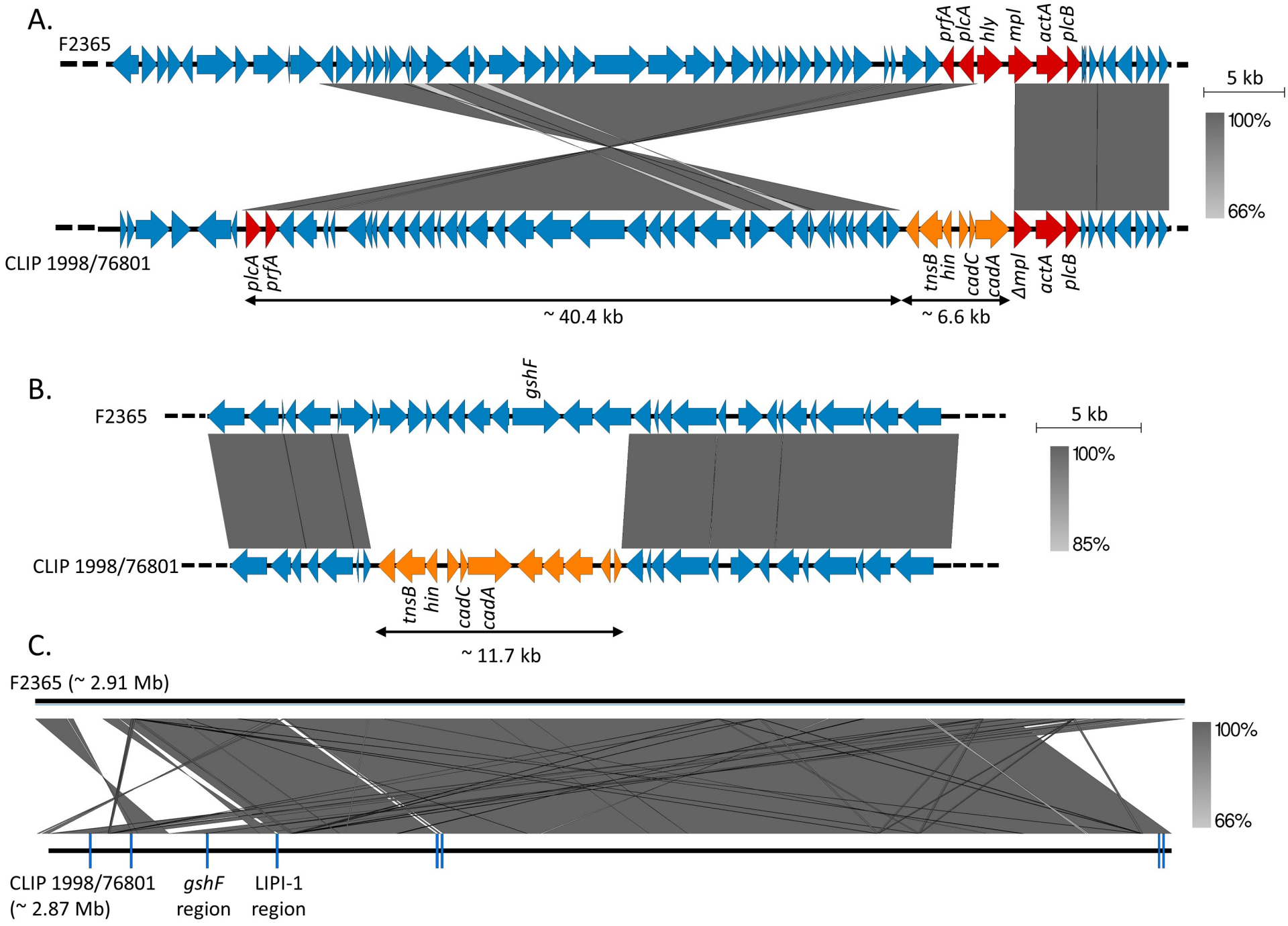
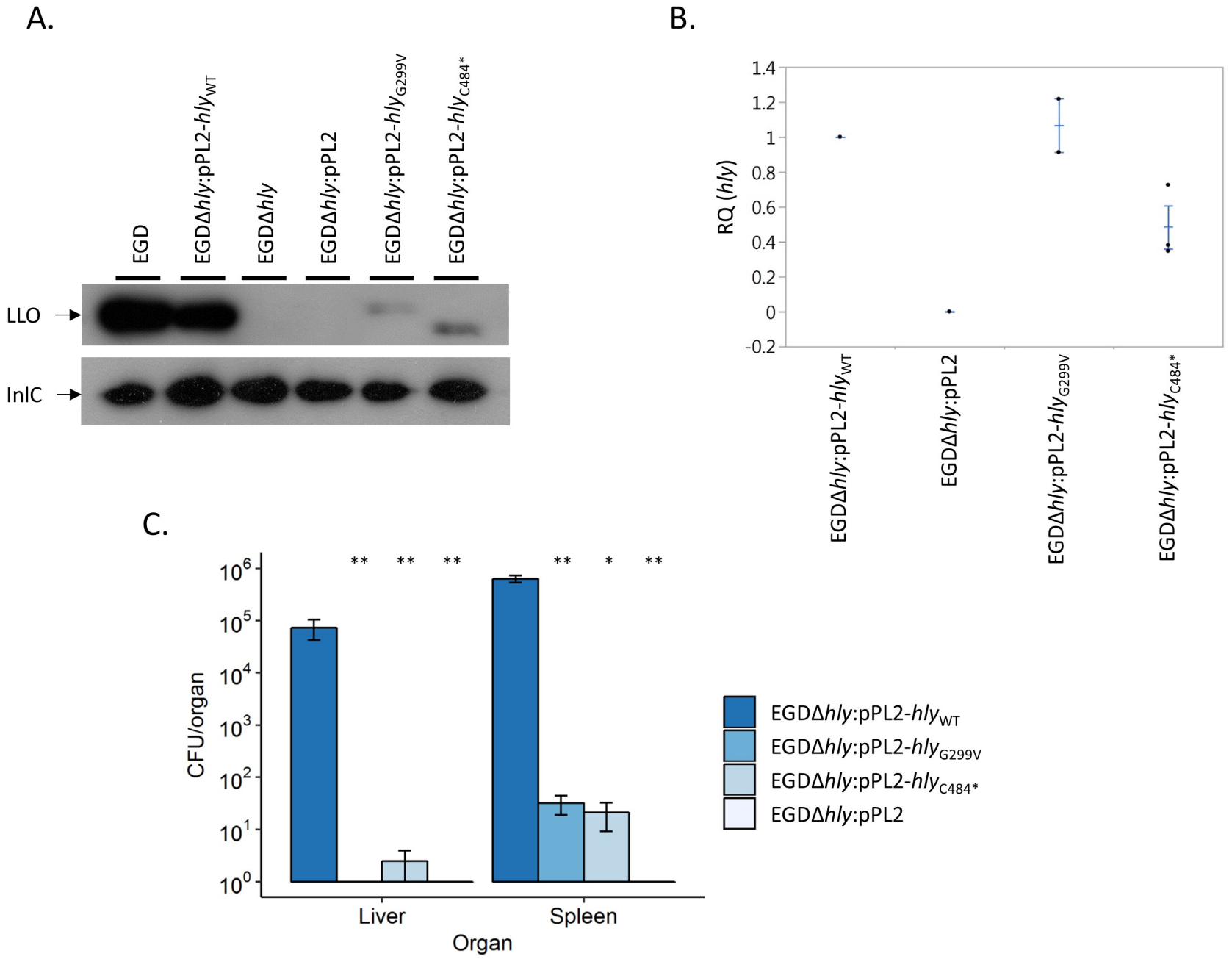

A.
F2365
prfA
plcA
hly
mpl
actA
plcB
5 kb
100%
66%
CLIP 1998/76801
plcA
prfA
tnsB
hin
cadC
cadA
Δmpl
actA
plcB
~ 40.4 kb
~ 6.6 kb
B.
gshF
F2365
5 kb
100%
85%
CLIP 1998/76801
tnsB
hin
cadC
cadA
~ 11.7 kb
C.
F2365 (~ 2.91 Mb)
100%
66%
CLIP 1998/76801
(~ 2.87 Mb)
gshF
region
LIPI-1
region

Figure 3


A.
EGD
EGDΔhly:pPL2-hly_WT
EGDΔhly
EGDΔhly:pPL2
EGDΔhly:pPL2-hly_G299V
EGDΔhly:pPL2-hly_C484*
LLO
InlC
B.
RQ (hly)
1.4
1.2
1
0.8
0.6
0.4
0.2
0
-0.2
EGDΔhly:pPL2-hly_WT
EGDΔhly:pPL2
EGDΔhly:pPL2-hly_G299V
EGDΔhly:pPL2-hly_C484*
C.
CFU/organ
10^6
10^5
10^4
10^3
10^2
10^1
10^0
**
**
**
**
*
**
Liver
Spleen
Organ
EGDΔhly:pPL2-hly_WT
EGDΔhly:pPL2-hly_G299V
EGDΔhly:pPL2-hly_C484*
EGDΔhly:pPL2